FEBS Letters 585 (2011) 2979–2985

journal homepage: www.FEBSLetters.org

Sobemovirus RNA linked to VPg over a threonine residue

Allan Olspert [a], Liisa Arike [b], Lauri Peil [b,c,d], Erkki Truve [a,*]

[a] Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia
[b] Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia
[c] Estonian Biocentre, Riia 23b, 51010 Tartu, Estonia
[d] Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

ARTICLE INFO

Article history:
Received 30 June 2011
Revised 3 August 2011
Accepted 4 August 2011
Available online 17 August 2011

Edited by Tamas Dalmay

Keywords:
Southern bean mosaic virus
Ryegrass mottle virus
RNA linkage
Phosphorylation
Mass spectrometry
Viral protein genome linked
Protein RNA covalent bond

ABSTRACT

Positive sense ssRNA virus genomes from several genera have a viral protein genome-linked (VPg) attached over a phosphodiester bond to the 5′ end of the genome. The VPgs of *Southern bean mosaic virus* (SBMV) and *Ryegrass mottle virus* (RGMoV) were purified from virions and analyzed by mass spectrometry. SBMV VPg was determined to be linked to RNA through a threonine residue at position one, whereas RGMoV VPg was linked to RNA through a serine also at the first position. In addition, we identified the termini of the corresponding VPgs and discovered three and seven phosphorylation sites in SBMV and RGMoV VPgs, respectively. This is the first report on the use of threonine for linking RNA to VPg.

1. Introduction

The 5′ ends of single-stranded positive-sense RNA virus genomes are unmodified, capped or have a viral protein genome-linked (VPg). The VPgs are attached to the RNA over a phosphodiester bond formed between the 5′ phosphate group of RNA and the hydroxyl group of an amino acid (aa) residue usually situated near the N-terminus of the protein. The aa residues involved in the VPg–RNA linkage have been reported to be either tyrosine or serine [1,2]. Picornaviruses utilize a highly conserved tyrosine residue situated near the N-terminus of VPg [1,3,4]. The use of tyrosine has also been shown for poty- and caliciviruses [5–7], whereas RNA linkage through a serine residue has been demonstrated for nepo- and comoviruses [2,8]. Recently we demonstrated that, dependent on the virus, sobemoviruses can utilize either tyrosine or serine [9]. Although threonine also contains a hydroxyl group, its use for linking RNA to VPg has not been reported.

Phosphodiester bonds exist also between DNA and proteins. Terminal proteins (TP) serve as primers for the synthesis of genomes and subsequent protective agents of genome termini of DNA viruses, mitochondrial plasmids and linear chromosomes [10,11]. Bond formation between TPs and DNA has been demonstrated to occur over the hydroxyl group of either serine, tyrosine or in that case, also threonine residues [10].

The VPgs of picornaviruses have been established mainly as primers for RNA synthesis [12], whereas potyviral VPgs are involved in the inhibition of cap-dependent and enhancement of cap-independent viral RNA translation [13].

Southern bean mosaic virus (SBMV) and Ryegrass mottle virus (RGMoV) are members of the genus Sobemovirus (reviewed in [14], genome organization revised in [15]), a group of small spherical viruses with a positive-sense single stranded RNA genome of approximately 4–4.5 kb. In sobemoviruses, the VPg is cleaved from the polyprotein by the viral protease [9,16–19] and is covalently attached to the 5′ end of genomic and subgenomic RNAs [20,21].

Sobemoviral VPgs are not conserved and unique in sense of RNA linking as they use either tyrosine or serine residues [9]. All this makes it difficult to predict the aa residue responsible for RNA linking for each virus species. The predicted RGMoV VPg does not have a suitable aa residue for linking near the N-terminus, whereas for SBMV VPg sequence comparisons suggest threonine as the most likely candidate [9]. We analyzed the VPg–RNA linkage of these two sobemoviruses. As a result we identified the true cleavage sites of RGMoV VPg, demonstrated that threonine is indeed used for linking RNA to SBMV VPg and described the post-translational modifications of these two VPgs.

* Corresponding author. Fax: +372 6204401.
E-mail address: erkki.truve@ttu.ee (E. Truve).

0014-5793/$36.00
doi:10.1016/j.febslet.2011.08.009

2. Materials and methods

For the infection, 2 week old oat (*Avena sativa* cv. Jaak) and bean (*Phaseolus vulgaris* cv. Sonesta or Aura) plants were mechanically inoculated with RGMoV (Japanese isolate, PV-307043 obtained from MAFF GenBank) and SBMV (Colombian isolate, PV-0100 obtained from DSMZ), respectively. After 4–5 weeks the leaves were harvested, virus particles and VPg were purified as described [9]. Briefly, virions were purified by ultracentrifugation, dissociated and the RNA was isolated. VPg, covalently bound to the RNA, was trypsin digested and subsequently the RNA was hydrolyzed in 10% trifluoroacetic acid for 48 h at room temperature. The samples were then dried under vacuum, purified with StageTips [22] and analyzed by LC–MS/MS using an Agilent 1200 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap mass-spectrometer (Thermo Electron) equipped with a nanoelectrospray ion source (Proxeon), as described before [9].

LTQ Orbitrap was operated in the data dependent mode with a full scan in the Orbitrap (mass range m/z 300–1900, resolution 60 000 at m/z 400, target value 1×10^6 ions) followed by up to five MS/MS scans in the LTQ part of the instrument (normalized collision energy 35%, wideband activation enabled, target value 5000 ions). Fragment MS/MS spectra from raw files were extracted as MSM files and then merged to peak lists using Raw2MSM version 1.11, selecting top eight peaks for each 100 Da [23]. MSM files were searched with the Mascot 2.3 search engine (Matrix Science) against the protein sequence database composed of VPg sequences and common contaminant proteins such as trypsin, keratins etc. Search parameters were as follows: 5 ppm precursor mass tolerance and 0.6 Da MS/MS mass tolerance, three missed trypsin cleavages plus a number of variable modifications such as oxidation (M), oxidation (HW), ethyl (DE), phospho (ST), phospho (Y), pAp (SYT), pGp (SYT), pCp (SYT) and pUp (SYT). For both viruses at least two independent biological samples were analyzed, each biological sample was in turn analyzed twice.

3. Results

3.1. Characterization of SBMV and RGMoV VPgs

VPgs purified from SBMV and RGMoV virions were trypsin-digested and studied with tandem MS analysis. The sequence coverage of SBMV VPg was 81–84%, identified peptides are shown in Table 1 and Fig. 1. We confirmed that the SBMV VPg is 77 aa residues in length, spanning from residue 326 to 402 in polyprotein P2a (Fig. 1A). Peptides between residues 30–38 and 65–69 were not detected, most likely due to their small size. Both N-terminal and C-terminal SBMV VPg protease cleavage sites were found to be between E/T residues.

The N-terminal peptide of RGMoV VPg was not detected when using the predicted annotated sequence as reference (accession NP_736586). Since sobemoviral proteases are known to cleave between E/T, E/S and E/N residues [16–18], we extended the N-terminal sequence to possible cleavage sites further upstream and discovered that the RGMoV VPg N-terminus is cleaved between E/S, three aa residues upstream of the previously proposed E/N site (Fig. 2A). The sequence coverage of RGMoV VPg was 91–94% and the length of the VPg was 79 aa residues. Identified peptides are shown in Table 1 and Fig. 2. The RGMoV VPg is cleaved from the P2a polyprotein between E/S residues at positions 314/315 and 393/394. Peptides between aa residue positions 33–36 of the RGMoV VPg were not detected in our study, similarly to short peptides from SBMV VPg.

3.2. SBMV and RGMoV VPg post-translational modifications

When searching for VPg–RNA linkage sites, we utilized the knowledge that the corresponding aa residue modification after RNA hydrolysis is a 5′,3′-diphosphate nucleotide, pNp (N denoting adenosine, cytidine, guanosine or uridine) [9], and expanded all these possible modifications to all possible phosphodiester bond acceptor residues (serine, tyrosine and threonine). By this approach we determined that the SBMV VPg was linked to RNA through a threonine residue at position one and the corresponding modification was pAp as assigned by modification delta mass and corresponding fragmentation spectrum (Fig. 1B). The RGMoV VPg was determined to be linked to RNA through a serine residue at position one and the modification was again assigned to be pAp (Fig. 2B). For both viruses the VPg N-terminal peptide was never detected in Mascot database search without the nucleotide modification.

In addition we identified several phosphorylation sites in both SBMV and RGMoV VPg-s (Table 1). In the SBMV VPg, serines at positions 7, 20 and 58 were found to be phosphorylated

Table 1
Examples of detected peptides identified by fragmentation spectra. The post-translational modifications are described and the modified position is in bold in the peptide sequence.

Virus	Position	Peptide	Modification	Experimental mass (Da)	Calculated mass (Da)	Mascot score
SBMV	1–19	**T**LPPELSVIEIPFEDVETR + pAp	pAp	2592.16	2592.16	53
SBMV	1–19	**T**LPPEL**S**VI**E**IPFEDVETR + pAp + P	pAp + phosphorylation + ethylation	2700.17	2700.15	60
SBMV	20–29	SYEFIEVEIK		1255.63	1255.63	72
SBMV	20–29	**S**YEFIEVEIK + P	Phosphorylation	1335.6	1335.6	65
SBMV	39–49	REFAWIPESGK		1318.67	1318.67	47
SBMV	40–49	EFAWIPESGK		1162.57	1162.57	72
SBMV	50–64	YWADDDDDSLPPPPK		1729.75	1729.75	99
SBMV	50–64	YWADDDDD**S**LPPPPK + P	Phosphorylation	1809.71	1809.71	99
SBMV	70–77	MVWSSAQE		936.4	936.4	42
RGMoV	1–10	**S**SENGEQGAR + pAp	pAp	1442.46	1442.46	28
RGMoV	1–10	**SS**ENGEQGAR + pAp + P	pAp (S1) + phos.(S2)	1522.43	1522.43	19
RGMoV	11–20	EIDAEEWISR		1246.58	1246.58	79
RGMoV	11–20	EIDAEEWI**S**R + P	Phosphorylation	1326.55	1326.55	42
RGMoV	21–32	EIDAEEWISREV**T**PTDVYIAGR + P	Phosphorylation	2628.22	2628.21	65
RGMoV	21–32	EIDAEEWI**S**REV**T**PTDVYIAGR + 2xP	Phosphorylation	2708.18	2708.18	54
RGMoV	21–32	EVTPTDVYIAGR		1319.68	1319.67	88
RGMoV	21–32	EV**T**PTDVYIAGR + P	Phosphorylation	1399.64	1399.64	56
RGMoV	37–54	VAGDEFSHSSYDPLAFSK		1955.9	1955.89	68
RGMoV	55–59	YKKER		722.41	722.41	22
RGMoV	58–79	ERGEMTWADMVEGDLDWDAREE		2639.09	2639.09	71
RGMoV	60–79	GEMTWADMVEGDLDWDAREE		2353.95	2353.95	91
RGMoV	60–79	GEM**T**WA**D**MVEGDLDWDAREE + P	Phosphorylation + ethylation	2461.94	2461.94	45

(Fig. 1B–D). In the RGMoV VPg on the other hand, we found phosphorylations on multiple serines, threonines and also on one tyrosine residue (Table 1). Of these listed modification sites, only phosphorylations at positions Ser2, Ser19, Thr23 and Thr63 were determined unambiguously (Fig. 2B–E). Serines at positions 43, 45 and 46 were identified by Mascot to be phosphorylated, but since the same MS2 spectrum had similar high ion scores for each of the potential phosphorylation sites (Table 2, Supplementary data), exact modification site assignments remain ambiguous. One of the possible phosphorylated serine assignments, Ser45 (best-scoring for this scan) is shown in Fig. 3A in bold, with other potentially phosphorylated serines underlined. While it is possible that all of these serine residues are phosphorylated, this result can also be explained by the known phenomenon of gas-phase rearrangements of phosphate groups. Namely, it has been shown that phosphate groups can be transferred to neighboring unmodified hydroxyl-containing amino acid residues upon collisional induced dissociation in linear ion traps, rendering many of the phosphorylation site assignments uncertain [24].

Interestingly, when the same peptide was detected as a doubly phosphorylated peptide, one of the site assignments, Tyr47, became unambiguous throughout Mascot searches (Fig. 2F, Table 2, Supplementary data), whereas it was still not possible to pinpoint confidently which of the serines was phosphorylated (Fig. 3B, Table 2, Supplementary data). In Fig. 3B, unambiguous phospho-Tyr47 is displayed in bold, best-scoring phospho-serine (Ser43) for this scan in underlined italic and other potentially phosphorylated serines are underlined. Therefore, we can conclude that Tyr47 is phosphorylated in combination with any of the three neighboring serines (Ser43, Ser45 or Ser46) being phosphorylated at the same time. For both SBMV and RGMoV, all phosphorylated peptides were also detected without the phosphorylations.

In addition to biologically relevant modifications, we detected random aspartate and/or glutamate ethylation(s), together with methionine and tryptophan oxidations (data not shown). These modifications are known to be generated in vitro during sample preparation [25,26] and were therefore not considered to be of biological relevance. All the VPg peptides usually detected in course of

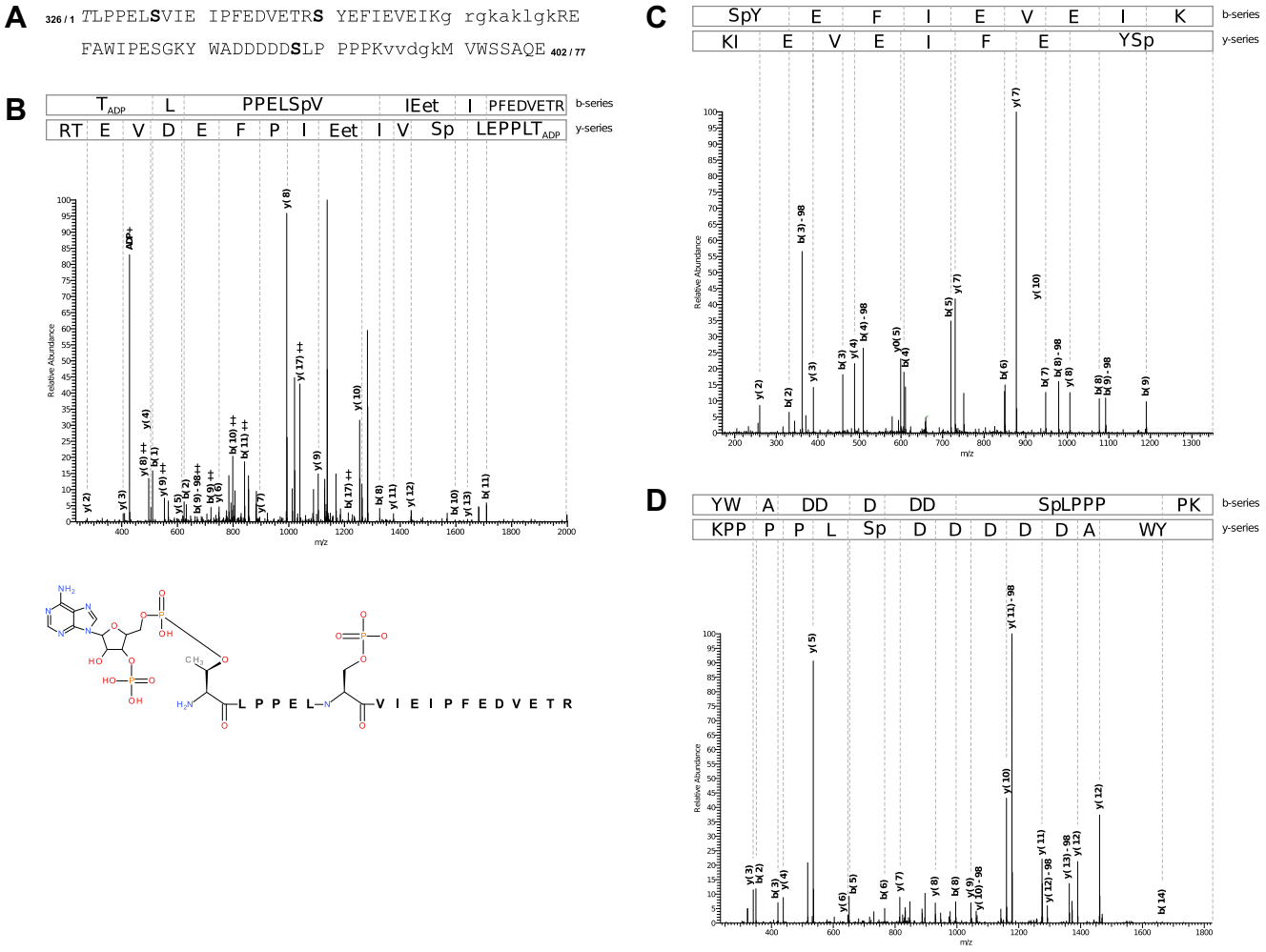

Fig. 1. Mass-spectrometrical characterization of SBMV VPg. (A) Representation of SBMV VPg. Numeration, indicated in subscript, corresponds to P2a polyprotein and VPg, respectively. The amino acid residue linked to RNA is in italics, phosphorylated residues are in bold and the region not detected is in lowercase. (B–D) Identification of post-translational modifications of SBMV VPg by MS/MS analysis. Co-purified VPg linked to RNA was trypsin-digested and RNA was degraded with acidic hydrolysis. The peptides were analyzed by nano-LC/MS/MS and resulting data was searched against corresponding sequence databases by MASCOT. The b and y ions represent peptide N- and C-terminal fragment ions produced by collision-induced dissociation in the mass spectrometer. (B) Determination of the residue covalently linked to RNA. The N-terminal SBMV VPg peptide, *T*LPPEL**S**VIE IPFEDVETR, was determined to contain a T1 linked pAp modification, a corresponding degradation product of viral RNA, and an additional phosphorylation at S6. The peptide sequence with the modification structure is represented below fragmentation spectrum. (C–D) Peptides **S**YEFIEVEIK and YWADDDDD**S**LP PPPK were detected to contain phosphorylations at respective positions.

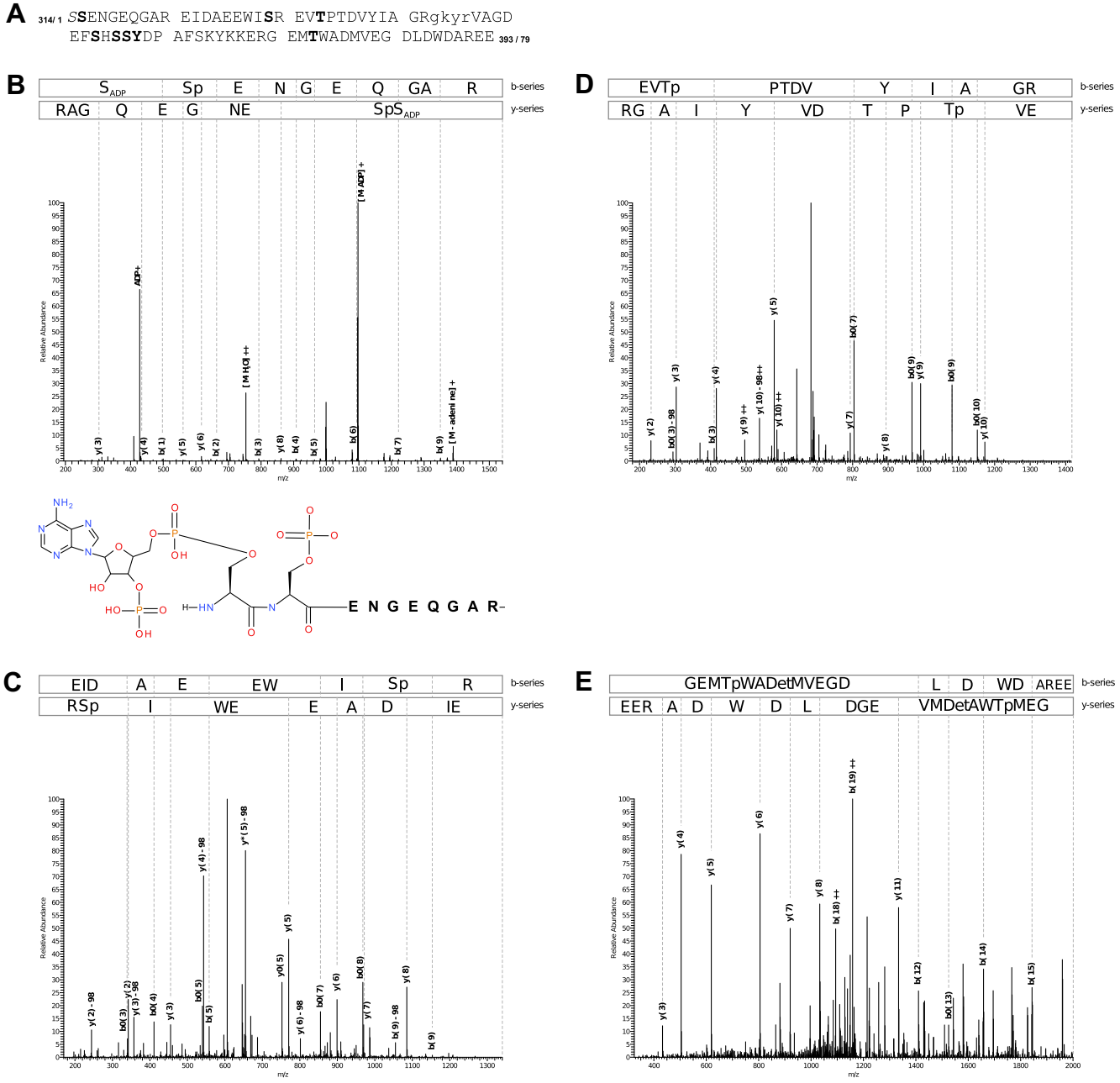

Fig. 2. Mass-spectrometrical characterization of RGMoV VPg. (A) Representation of RGMoV VPg. Numeration, indicated in subscript, corresponds to P2a polyprotein and VPg, respectively. The amino acid residue linked to RNA is in italics, phosphorylated residues are in bold and the region not detected is in lowercase. (B–E) Identification of post-translational modifications of RGMoV VPg by MS/MS analysis. Co-purified VPg linked to RNA was trypsin-digested and RNA was degraded with acidic hydrolysis. The peptides were analyzed by nano-LC/MS/MS and resulting data was searched against corresponding sequence databases by MASCOT. The b and y ions represent peptide N- and C-terminal fragment ions produced by collision-induced dissociation in the mass spectrometer. (B) The N-terminal RGMoV VPg peptide, S**S**ENGEQGAR, was determined to contain a S1 linked pAp modification and an additional phosphorylation at S2. The peptide sequence with the modification structure is represented below fragmentation spectrum. (C–E) Peptides EIDAEEWI**S**R, EV**T**PTDVYIAGR and GEM**T**WADMVEGDLDWDAR were detected to contain phosphorylations at respective positions. Peptide GEM**T**WADMVEGDLDWDAR also contained a in vitro induced ethylation (et) at position D6.

these analyses (except for oxidized and ethylated forms) are shown in Tables 1–2.

4. Discussion

The VPg aa residue reported to be linked to RNA has so far been either serine or tyrosine. In addition to these two aa residues, threonine also contains a hydroxyl group and has been shown to be used for phosphodiester bond formation with DNA [10]. Furthermore, there are reports about the use of threonine in RNA linking of cellular non-viral proteins [27]. Sobemoviral VPg sequences are not conserved at genus level and we recently demonstrated that individual sobemoviruses can somewhat surprisingly use different aa residues at different positions for RNA linking [9]. Sequence comparisons of sobemoviral VPgs as well as experimental data suggested that sobemoviruses might also use threonine for RNA linking as the threonine in position 1 of SBMV VPg was the only residue out of the first 20 which was not detected correctly by Edman sequencing [16]. Here we report for the first time that threonine can indeed be utilized for linking RNA to VPg. This

Table 2
Overview of RGMoV VPg peptide VAGDEFSHSSYDPAFSK phosphorylation annotations. For each listed MS2 scan, up to five highest-scoring Mascot assignments are shown. For singly phosphorylated peptide, the exact modification site is uncertain between different scans whereas for doubly phosphorylated peptide Tyr47 is always the best-scoring one. Full Mascot results for individual scans are shown in detail in the Supplementary data.

Virus	Position	Experiment/ injection	Scan number	Peptide	Modification	Experimental mass (Da)	Calculated mass (Da)	Mass error (ppm)	Mascot score
RGMoV	37–54	1/1	5766	VAGDEFSHSSYDPLAFSK	S9 phosphorylation	2035.8583	2035.8564	0.92	47
					S7 phosphorylation				47
					S10 phosphorylation				44
					Y11 phosphorylation				37
RGMoV	37–54	1/2	5580	VAGDEFSHSSYDPLAFSK	S7 phosphorylation	2035.8597	2035.8564	1.6	32
					S9 phosphorylation				26
					S10 phosphorylation				21
					Y11 phosphorylation				16
RGMoV	37–54	1/1	6141	VAGDEFSHSSYDPLAFSK	S9 phosphorylation	2035.86	2035.8564	1.75	62
					S10 phosphorylation				61
					S7 phosphorylation				51
					Y11 phosphorylation				46
					S17 phosphorylation				40
RGMoV	37–54	1/2	5911	VAGDEFSHSSYDPLAFSK	S10 phosphorylation	2035.8607	2035.8564	2.11	58
					S7 phosphorylation				56
					Y11 phosphorylation				50
					S9 phosphorylation				48
RGMoV	37–54	1/1	6259	VAGDEFSHSSYDPLAFSK	S9 + Y11 phosphorylation	2115.8269	2115.8228	1.94	62
					S7 + Y11 phosphorylation				62
					S10 + Y11 phosphorylation				62
					S9 + S10 phosphorylation				55
					S7 + S10 phosphorylation				53
RGMoV	37–54	1/2	6010	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8278	2115.8228	2.36	36
					S9 + Y11 phosphorylation				33
					S10 + Y11 phosphorylation				27
					S7 + S10 phosphorylation				27
					S9 + S10 phosphorylation				24
RGMoV	37–54	1/2	6011	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8295	2115.8228	3.2	43
					S7 + S10 phosphorylation				41
					S7 + S17 phosphorylation				35
					S10 + Y11 phosphorylation				34
					S9 + Y11 phosphorylation				34
RGMoV	37–54	2/2	5791	VAGDEFSHSSYDPLAFSK	S10 phosphorylation	2035.8554	2035.8564	−0.51	54
					S9 phosphorylation				47
					Y11 phosphorylation				41
					S7 phosphorylation				37
RGMoV	37–54	2/2	5797	VAGDEFSHSSYDPLAFSK	S7 phosphorylation	2035.8566	2035.8564	0.075	47
					S9 phosphorylation				39
					S10 phosphorylation				36
RGMoV	37–54	2/1	5851	VAGDEFSHSSYDPLAFSK	S10 phosphorylation	2035.8586	2035.8564	1.05	58
					S9 phosphorylation				56
					S7 phosphorylation				50
					S17 phosphorylation				45
					Y11 phosphorylation				40
RGMoV	37–54	2/2	5889	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8262	2115.8228	1.63	26
					S7 + S10 phosphorylation				22
					S9 + Y11 phosphorylation				20
					S7 + S9 phosphorylation				19
					S10 + Y11 phosphorylation				19
RGMoV	37–54	2/2	5892	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation	2115.8287	2115.8228	2.81	45
					S7 + S10 phosphorylation				35
					S7 + S9 phosphorylation				35
					S10 + Y11 phosphorylation				35
					S9 + Y11 phosphorylation				35
RGMoV	37–54	2-1	6086	VAGDEFSHSSYDPLAFSK	S7 + Y11 phosphorylation, D4 ethylation	2143.8579	2143.8541	1.78	49
					S9 + Y11 phosphorylation, D4 ethylation				42
					S7 + Y11 phosphorylation, E5 ethylation				41
					S9 + Y11 phosphorylation, E5 ethylation				40

demonstrates that sobemoviruses are even more diverse than expected and can use all available hydroxyl group containing aa residues for RNA linking. Furthermore, this finding also shows that threonine as an RNA linking residue cannot be ruled out for viruses from other genera. It is surprising that the use of threonine had not been described before, indicating that sobemoviruses are either truly unique in their variability or that the VPg linking as such has not been studied intensively enough. In addition to SBMV, a few VPgs of other sobemoviruses also have a threonine at their N-terminus – *Sowbane mosaic virus*, *Subterranean clover mottle virus*, SeMV and *Southern cowpea mosaic virus* (SCPMV). The latter two are very closely related to SBMV and their VPgs are conserved (65% identity and 79% similarity within the VPgs of the three viruses). With the exception of *Cocksfoot mottle virus* (CfMV) for which tyrosine at position five is used for RNA linking, the rest of sequenced sobemoviruses all have a serine residue at the putative first position of VPg. Furthermore, for SBMV, RGMoV and RYMV it has now been experimentally proven that it is the first VPg residue

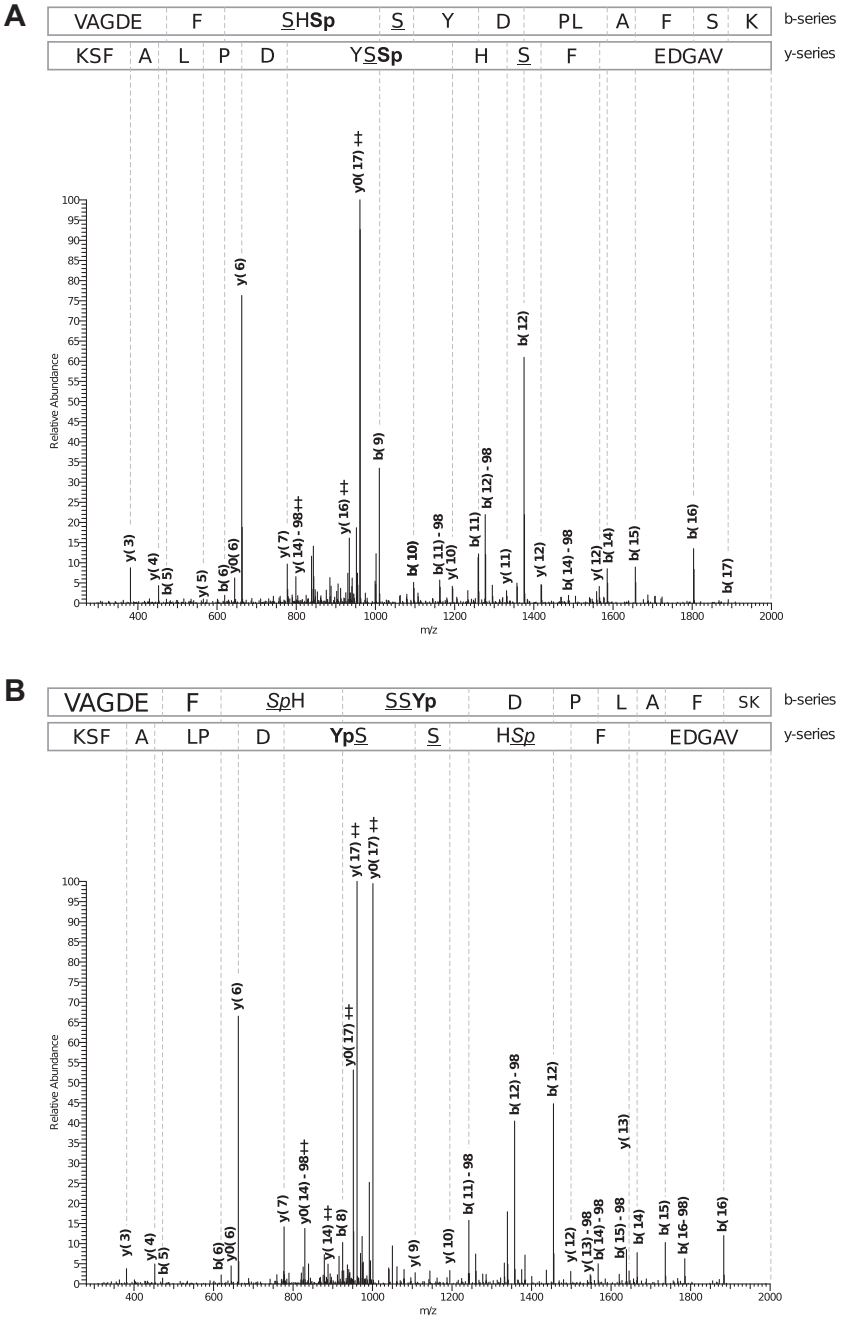

Fig. 3. Phosphorylation site assignments for RGMoV VPg peptide VAGDEFSHSSYDPAFSK. (A) Annotated fragmentation spectrum for singly phosphorylated peptide. Best-scoring phosphorylated serine assignment (for this scan) is shown in bold, other potentially phosphorylated positions are underlined. (B) Annotated fragmentation spectrum for doubly phosphorylated peptide. Unambiguous phospho-Tyr47 is displayed in bold, best-scoring phospho-serine (Ser43) for this scan in underlined italic and other potentially phosphorylated positions are underlined.

that is indeed used for RNA linking. Therefore, we conclude that threonine and serine residues at the first position of the VPgs of these viruses are most probably used for RNA linking.

The remainder of RNA degradation attached to the SBMV VPg was detected to be pAp, establishing that the first nucleotide of SBMV genomic RNA must be A. Intriguingly, all available SBMV (and SCPMV) genome sequences have the 5′ sequence starting with CACAAA. However, it has been demonstrated that at least for some isolates of SBMV and SCPMV the 5′ terminal C is erroneous and that the true 5′ sequence is ACAAA [28]. Our results independently confirm that the first nucleotide of the SBMV genome should indeed be A. Since CACAAA is reported to be the beginning of the genome of another very closely related species, SeMV, it would be interesting to see whether these viruses actually vary at the 5′ end of the genome or not.

We also confirmed that the 5′ sequence of RGMoV genome starts with an A nucleotide [29]. At the same time we discovered that putative N-terminus of RGMoV VPg was annotated incorrectly – mature RGMoV VPg has additional three amino acid residues, SSE, in its N-terminus. This discovery provides us with further information about the protease specificity of sobemoviruses. The experimentally proven protease cleavage sites for all sobemoviruses are E/T and E/S, with the only exception being CfMV that utilizes also E/N site for the cleavage of its VPg. For SeMV the proposed

specificity of the protease is (N or Q)–E/(T or S)–X, where X is an aliphatic residue [18]. While this proposed consensus applies to SeMV, to one CfMV site and also to SBMV (closely related to SeMV), it certainly does not apply for the whole genera.

Previously the phosphorylation of potyviral and sobemoviral VPg has been reported [9,30,31]. Comparing the phosphorylation sites of VPgs of CfMV, RYMV, SBMV and RGMoV only indicates that the protein sequences as well as their phosphorylation patterns are diverse. However, based on VPg sequence similarity, the data on SBMV allows us to make predictions about the phosphorylation of SeMV and SCPMV VPgs. Serines 7 and 20 are present at the same positions in SeMV and serine 20 in SCPMV, which indicates that these residues might also be the targets for phosphorylation. The position corresponding to serine 58 of SBMV VPg is occupied by glutamic acid in SeMV VPg and by aspartic acid in SCPMV VPg, both of them chemically mimic phosphoserine to some extent. This indicates that the negative charge at VPg position 58 might be important for all three viruses. Altogether, sobemoviral VPgs, which are anyhow rich in negatively charged aa residues seem to require an additional overall negative charge, achieved through multiple phosphorylations. In contrast, the N-termini of sobemoviral coat proteins are rich in positively charged aa residues [14] and reside inside the virion, presumably in contact with negatively charged RNA and VPg [32–34]. Therefore the phosphorylation of VPg might be needed for providing additional stability to the virion through electrostatic interactions. Nevertheless, one also cannot rule out other possible functions of VPg phosphorylation throughout the virus infection cycle. For instance, in the in vitro assay RdRp of SeMV failed to nucleotidylate VPg purified from bacteria where presumably phosphorylation does not occur similarly to plant cells [35]. The authors propose that the nucleotidylation reaction could be dependent on (host) factors missing in the in vitro assay. VPg phosphorylation could be one factor regulating RdRp VPg interactions required for this process.

Acknowledgments

We wish to thank Signe Nõu for excellent plant care. This work was supported by Estonian Science Foundation grant no. 7363. Mass spectrometric analyses were in part supported by the European Regional Development Fund through the Center of Excellence in Chemical Biology (Institute of Technology, University of Tartu) and by the ECOGENE project (EC grant number 205419 to Estonian Biocentre).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2011.08.009.

References

[1] Ambros, V. and Baltimore, D. (1978) Protein is linked to the 5′ end of poliovirus RNA by a phosphodiester linkage to tyrosine. J. Biol. Chem. 253, 5263–5266.
[2] Jaegle, M., Wellink, J. and Goldbach, R. (1987) The genome-linked protein of *Cowpea mosaic virus* is bound to the 5′ terminus of virus RNA by phosphodiester linkage to serine. J. Gen. Virol. 68, 627–632.
[3] Rothberg, P.G., Harris, T.J., Nomoto, A. and Wimmer, E. (1978) O4-(5′-uridylyl)tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. Proc. Natl. Acad. Sci. U.S.A. 75, 4868–4872.
[4] Schein, C.H., Oezguen, N., Volk, D.E., Garimella, R., Paul, A. and Braun, W. (2006) NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. Peptides 27, 1676–1684.
[5] Anindya, R., Chittori, S. and Savithri, H.S. (2005) Tyrosine 66 of *Pepper vein banding virus* genome-linked protein is uridylylated by RNA-dependent RNA polymerase. Virology 336, 154–162.
[6] Belliot, G., Sosnovtsev, S.V., Chang, K.O., McPhie, P. and Green, K.Y. (2008) Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. Virology 374, 33–49.
[7] Murphy, J.F., Rychlik, W., Rhoads, R.E., Hunt, A.G. and Shaw, J.G. (1991) A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. J. Virol. 65, 511–513.
[8] Zalloua, P.A., Buzayan, J.M. and Bruening, G. (1996) Chemical cleavage of 5′-linked protein from tobacco ringspot virus genomic RNAs and characterization of the protein–RNA linkage. Virology 219, 1–8.
[9] Olspert, A., Peil, L., Hébrard, E., Fargette, D. and Truve, E. (2011) Protein–RNA linkage and post-translational modifications of two sobemovirus VPgs. J. Gen. Virol. 92, 445–452.
[10] Salas, M. (1991) Protein-priming of DNA replication. Annu. Rev. Biochem. 60, 39–71.
[11] Fricova, D., Valach, M., Farkas, Z., Pfeiffer, I., Kucsera, J., Tomaska, L. and Nosek, J. (2010) The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5′ termini. Microbiology 156, 2153–2163.
[12] Steil, B.P. and Barton, D.J. (2009) Conversion of VPg into VPg-pUpU-OH before and during poliovirus negative-strand RNA synthesis. J. Virol. 83, 12660–12670.
[13] Khan, M.A., Miyoshi, H., Gallie, D.R. and Goss, D.J. (2008) Potyvirus genome-linked protein, VPg, directly affects wheat germ in vitro translation: interactions with translation initiation factors eIF4F and eIFiso4F. J. Biol. Chem. 283, 1340–1349.
[14] Tamm, T. and Truve, E. (2000) Sobemoviruses. J. Virol. 74, 6231–6241.
[15] Meier, M. and Truve, E. (2007) Sobemoviruses possess a common CfMV-like genomic organization. Arch. Virol. 152, 635–640.
[16] van der Wilk, F., Verbeek, M., Dullemans, A. and van den Heuvel, J. (1998) The genome-linked protein (VPg) of southern bean mosaic virus is encoded by the ORF2. Virus Genes 17, 21–24.
[17] Mäkinen, K., Mäkeläinen, K., Arshava, N., Tamm, T., Merits, A., Truve, E., Zavriev, S. and Saarma, M. (2000) Characterization of VPg and the polyprotein processing of cocksfoot mottle virus (genus sobemovirus). J. Gen. Virol. 81, 2783–2789.
[18] Nair, S. and Savithri, H.S. (2010) Processing of SeMV polyproteins revisited. Virology 396, 106–117.
[19] Hébrard, E., Pinel-Galzi, A. and Fargette, D. (2008) Virulence domain of the RYMV genome-linked viral protein VPg towards rice rymv1–2-mediated resistance. Arch. Virol. 153, 1161–1164.
[20] Ghosh, A., Rutgers, T., Ke-Qiang, M. and Kaesberg, P. (1981) Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. J. Virol. 39, 87–92.
[21] Mang, K.Q., Ghosh, A. and Kaesberg, P. (1982) A comparative study of the cowpea and bean strains of southern bean mosaic virus. Virology 116, 264–274.
[22] Rappsilber, J., Mann, M. and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stagetips. Nat. Protoc. 2, 1896–1906.
[23] Olsen, J.V., de Godoy, L.M.F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. and Mann, M. (2005) Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a c-trap. Mol. Cell Proteomics 4, 2010–2021.
[24] Palumbo, A.M. and Reid, G.E. (2008) Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-ms/ms and ms3. Anal. Chem. 80, 9735–9747.
[25] Xing, G., Zhang, J., Chen, Y. and Zhao, Y. (2008) Identification of four novel types of *in vitro* protein modifications. J. Proteome Res. 7, 4603–4608.
[26] Stadtman, E.R. and Levine, R.L. (2003) Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. Amino Acids 25, 207–218.
[27] Thomas, L., Pfeifle, J. and Anderer, F.A. (1987) Human SS-b/La autoantigen contains a covalent protein–RNA linkage. Biochim. Biophys. Acta 909, 173–182.
[28] Hacker, D.L. and Sivakumaran, K. (1997) Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. Virology 234, 317–327.
[29] Balke, I., Resevica, G. and Zeltins, A. (2007) The ryegrass mottle virus genome codes for a sobemovirus 3c-like serine protease and RNA-dependent RNA polymerase translated via -1 ribosomal frameshifting. *Virus Genes*. 35, 395–398.
[30] Puustinen, P., Rajamäki, M., Ivanov, K.I., Valkonen, J.P.T. and Mäkinen, K. (2002) Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. J. Virol. 76, 12703–12711.
[31] Hafrén, A. and Mäkinen, K. (2008) Purification of viral genome-linked protein VPg from potato virus a-infected plants reveals several post-translationally modified forms of the protein. J. Gen. Virol. 89, 1509–1518.
[32] Plevka, P., Tars, K., Zeltins, A., Balke, I., Truve, E. and Liljas, L. (2007) The three-dimensional structure of ryegrass mottle virus at 2.9 Å resolution. Virology 369, 364–374.
[33] Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G., Rayment, I., Rossmann, M.G., Suck, D. and Tsukihara, T. (1980) Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 286, 33–39.
[34] Rossmann, M.G., Abad-Zapatero, C., Erickson, J.W. and Savithri, H.S. (1983) RNA–protein interactions in some small plant viruses. J. Biomol. Struct. Dyn. 1, 565–579.
[35] Govind, K. and Savithri, H.S. (2010) Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase. Virology 401, 280–292.